

Report by: Anand Prashar

NLP Homework 5 – q4

Q:

Now let's examine similarities in trained word embeddings, instead of raw context counts. See the file `Nyt_word2vec.university cat dog`, which contains word embedding vectors pretrained by word2vec₁ for three words: "dog", "cat", and "university". You can open the file in a text editor since it's quite small.

Please complete `cossim` and `dense(v1,v2)` in `distsim.py` to compute and display the cosine similarities between each pair of these words. You may not use any functions from packages (e.g. `scikit`) that contain implementations of cosine calculation. You can test that your implementation is correct with `testq4.py`. We will also test this on Vocareum.

A: Result of `testq4.py`

```
Cosine similarity between cat and dog 0.827517295965
Cosine similarity between cat and university -0.205394745036
Cosine similarity between university and dog -0.190753135501
1: cat (0.827517295965)
2: university (-0.190753135501)
```